

Variational Inference

From EM-days, we know, for "any" $q(z)$

$$\log P(x|\theta) \geq \underbrace{E_{q(z)} \log P(x, z|\theta)}_{\substack{\uparrow \text{can be derived with Jensen's} \\ \text{inequality}}} - E_{q(z)} \log q(z)$$

What is the difference b/w L.H.S & R.H.S ?

$$\log P(x|\theta) - E_{q(z)} \log P(x, z|\theta) + E_{q(z)} \log q(z)$$

\Downarrow same

$$E_{q(z)} \log P(x|\theta) - E_{q(z)} \log P(x, z|\theta) + E_{q(z)} \log q(z)$$

$$E_{q(z)} \log \frac{P(x|\theta)}{P(x, z|\theta)} + E_q \log q(z)$$

$$= - \left[E_{q(z)} \log \frac{P(x, z|\theta)}{P(x|\theta)} - E_q \log q(z) \right]$$

$$= - E_{q(z)} \log \frac{P(z|x, \theta)}{q(z)}$$

$$= KL(q, P(z|x, \theta)) \therefore \text{Kulback-Leibler divergence b/w } q \text{ and } P(z|x, \theta)$$

Hence

$$\log P(x|\theta) = \underbrace{E_{q(z)} \log \frac{P(x, z|\theta)}{q(z)}}_{\substack{\text{lower-bound} \\ L(q; \theta)}} - \underbrace{E_{q(z)} \log \frac{P(z|x, \theta)}{q(z)}}_{\text{KL}(q \| P(z|x, \theta))}$$

↑ function of q and θ

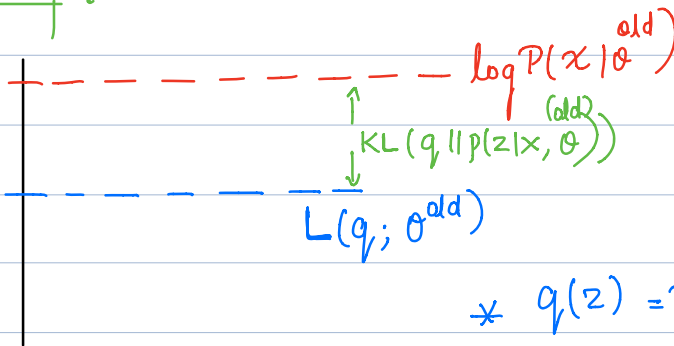
Alternative view of EM:-

* Coordinate ascent between $q(z)$ and θ

In E-step, $\theta = \theta^{(\text{old})}$ and we optimize $L(q; \theta^{(\text{old})})$
w.r.t $q(z)$, $q(z) \rightarrow p(z|x, \theta^{(\text{old})})$

In M-step, $q(z) = p(z|x, \theta^{(\text{old})})$ and we optimize
 $L(p(z|x, \theta^{(\text{old})}); \theta)$ w.r.t to θ

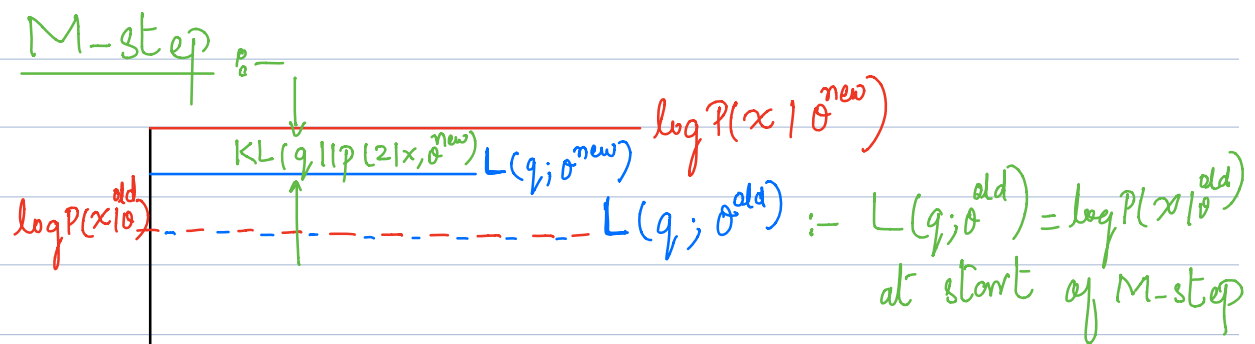
E-Step:-



* $q(z) = p(z|x, \theta^{(\text{old})})$ is the
solution of $\arg \max_{q(z)} L(q; \theta^{(\text{old})})$

* In E-step, $\log P(x | \theta^{\text{old}})$ does not improve emp
as independent of $q(z)$

* Only $KL(q || p(z|x, \theta)) \rightarrow 0$, when $q(z) \rightarrow p(z|x, \theta)$



* In M-step, $\theta^{\text{new}} \leftarrow \arg\max_{\theta} L(q(z)=p(z|x, \theta^{\text{old}}); \theta)$

* Also moves $\log P(x | \theta)$ up, and makes $KL(q || p(z|x, \theta^{\text{new}})) \geq 0$

* In a Bayesian model, $\theta \sim \text{Prior}(\theta | \text{hyper-param})$ ← like any latent variable

* We want $\Pr(\theta | x)$

* Essentially, we just want to run the E step above i.e.

$$\max_q L(q; \text{hyper-parameters})$$

(A)

where q is some distribution on latent variables

* $\max_q L(q; \text{hyperparameters})$ is optimization in the space of distributions.

* Variational inference deals with this optimization.

* If no constraint is put on the space of distributions, optimal $q = P(\text{latent variables} | x)$

* $P(\text{latent variables} | x)$ may not be possible to compute analytically. { for various reasons, we have discussed so far }

* Goal :- find a good easy approximation to $P(\text{latent variables} | x)$ by constraining space of "q"

* Two approaches :- a) assume a parametric form of "q" and optimize $L(q; \text{hyp})$ w.r.t parameters

b) Assume a factorized 'q' space. (mean-field approximation)

Any of the two approaches will (possibly) leave a non-zero $KL(q || P(\text{latent variables} | x))$

MEAN - FIELD APPROXIMATION

Let $Z = [z_1, \dots, z_N]$ all latent variables

$q(Z) = \prod_{i=1}^M q(z_i)$ is the constrained

state of distributions. Optimizing $L(q; \text{hyp})$

w.r.t such $q(Z)$ gives the following

$$\log q(z_i) = \underbrace{E_{\prod_{j:j \neq i} q(z_j)} \log P(x, z)}_{\text{expectation of } \log P(x, z) \text{ w.r.t } \prod_{j:j \neq i} q(z_j)} + \text{const}$$

if $\log = \log_e$ (natural log)

$$q(z_i) = \frac{\exp \bar{E}_{j:j \neq i} \ln P(x, z)}{\int \exp \bar{E}_{j:j \neq i} \ln P(x, z) dz_i} \quad \text{for } i=1, \dots, N$$

Proof:-

$$\begin{aligned}
L(q; \text{hyp}) &= \int \ln P(x, z | \text{hyp}) \prod_{i=1}^M q(z_i) dz_1 \dots dz_M \\
&\quad - \int \left[\ln \prod_{i=1}^M q(z_i) \right] \prod_{i=1}^M q(z_i) dz_1 \dots dz_M \\
&= \int q(z_i) \left[\overbrace{\int \ln P(x, z | \text{hyp}) \prod_{j:j \neq i} q(z_j) dz_j}^{E_{i \neq j} \ln P(x, z | \text{hyp})} \right] dz_i \\
&\quad - \left[\int (\ln q(z_i)) q(z_i) dz_i + \sum_{j \neq i} \int [\ln q(z_j)] q(z_j) dz_j \right]
\end{aligned}$$

for optimizing w.r.t $q(z_i)$ \nearrow Constant

$$\begin{aligned}
&= \int q(z_i) E_{j \neq i} \ln P(x, z) dz_i \\
&\quad - \int q(z_i) \ln q(z_i) dz_i
\end{aligned}$$

negative KL b/w $q(z_i)$ and $E_{i \neq j} \ln P(x, z | \text{hyp})$

$$\text{So } \ln q(z_i) = E_{j \neq i} \ln P(x, z) + \text{const}$$